The importance of recording and playback technique for assessment of annoyance

Celik, Emine; Persson Waye, Kerstin; Møller, Henrik

Published in:
Proceedings of Forum Acusticum

Publication date:
2005

The importance of recording and playback technique for assessment of annoyance

Emine Çelik
Department of Acoustics, DK 9220 Aalborg Ø, Fredrik Bajers Vej 7 B5, Denmark, emc@acoustics.aau.dk

Kerstin Persson Waye
Department of Environmental Medicine, S-405 30 Göteborg, Sweden, kerstin.persson-waye@envmed.gu.se

Henrik Møller
Department of Acoustics, DK 9220 Aalborg Ø, Fredrik Bajers Vej 7 B5, Denmark, hm@acoustics.aau.dk

The study presented here is part of a project with an overall aim to evaluate how various physical properties of sound relate to annoyance. In order to achieve this it is necessary to study methodological aspects of importance for the experimentally evaluated annoyance. In previous studies of perception and response to sounds, several methods have been adopted both with regard to recording techniques (monophonic or binaural), playback techniques (through headphones or loudspeakers) and subjective evaluation techniques. The present study was carried out to investigate if there is a difference in perception related to annoyance, loudness and unpleasantness between monophonic recordings played back through a loudspeaker and binaural recordings played back via headphones and to evaluate whether a possible difference depends on temporal, spectral and spatial characteristics of the sound. The experiment adopted two psychometric methods for achieving responses from subjects, and different durations of the exposure were used. Fifty-four subjects participated and three types of sounds were used in the experiments: everyday "restaurant" sounds (from using cutlery at platters, moving chairs, talking etc.), road traffic sound and a low-frequency ventilation sound. The sounds were recorded with two different techniques (monophonic and binaural) and each sound was played back at three different sound levels. The monophonic recordings were presented through a loudspeaker and the binaural recordings were presented through both closed (circum-aural) and completely open (free of the ear) headphones. The results show that for all judgments (annoyance, loudness and unpleasantness), there was no significant main effect of recording and playback techniques; however significant interactions between techniques and sounds were found.

1 Introduction

Experimental studies of noise annoyance are often criticized for using exposure conditions that are so unlike those of the real world that the results can not be applied to settings outside the laboratory. In order to ensure validity between real life and experimental settings, several conditions related to recording, playback and context of the experimental situation need to be attended to. In previous studies of perception and response to sounds, several methods have been adopted both with regard to recording techniques (monophonic or binaural), playback techniques (through headphones or loudspeakers) and subjective evaluation techniques. Regarding recording and playback techniques very little is known on how these techniques affect the subjective perception and overall response. A better knowledge in this field is crucial in order to compare sound exposures between studies. A major difference between the two recording and playback techniques is their ability to reproduce spatial properties of the sound. A further difference exists for low frequencies, which at higher sound pressure levels do not only affect the hearing but also give sensations in other parts of the body [2].

The purpose of this study is to investigate whether there is a difference in subjective perception and response related to annoyance, loudness and unpleasantness between mono recordings, played back through a loudspeaker, and binaural recordings played back via headphones. A further aim was to evaluate whether the perception differed depending on temporal, spectral and/or spatial characteristics of the sound. The study also adopts two psychometric methods for achieving responses from subjects. Many of the response methods used today are based on short-term comparisons of sounds and it can be questioned whether they can be used to measure annoyance or even unpleasantness. Therefore the project also aimed at evaluating the effect of exposure duration on the assessments. In the study, large efforts were undertaken to collect data that could be representative for real life, but in the same time control for errors caused by the recording and playback techniques, and the experimental room.

2 Material and methods

2.1 Subjects

A total of 54 paid native Danish speaking volunteers participated in the experiments (27 females and 27 males aged between 20 and 34 years, M=24.72, SD=2.78). The subjects had not previously participated in similar sound evaluation experiments. Audiometric tests (ISO 8253-1) ensured normal hearing within 15 dB at the octave band frequencies 125 Hz to 4 kHz and 20 dB at 8 kHz. To assess the subjects' noise sensitivity in general, a questionnaire [5] translated into Danish was answered after the audiometric tests. The questionnaire had a total of 120 points; the higher the point scores, the higher sensitivity to noise. The subjects' answers ranged between 48 and 111 points with an average of 72.5 (SD=11.83). The subjects were allocated to the three groups that judged different psychoacoustic attributes: annoyance, loudness and unpleasantness. Females and males were separately ordered on the basis of their noise sensitivity scores. The first three female subjects were randomly distributed into the three groups. The same was done to the first three male subjects, then to the next three female subjects etc. etc. This process went on until the end of the lists so that in the three groups there were equal numbers of female and male subjects having comparable noise sensitivity.

2.2 Sounds

Three sounds were used in the study. The sounds varied in particular with regard to spatial properties and content of low frequencies (20-200 Hz). The recordings were done with a Harmonie 01 dB system using an artificial head [1] for the binaural recordings and a G.R.A.S 40 EN microphone for the monophonic recordings. The first sound (R) comprised sounds typically occurring in a restaurant. Sounds from using cutlery at platters, moving chairs and people talking occurred in many directions. The conversations were done in Turkish (female voice) and Spanish (male voice in Costilla La Mancha accent) so the conversation would be meaningless to the test subjects. The second sound, traffic sound (T), was obtained from a road in front of the recording position and thus sound sources occurred in a limited spatial range in the original sound field. The third sound, ventilation sound (V), was recorded in a large basement room with ventilation channels, and there was no obvious direction to the sound source(s). In order to obtain a predominantly low frequency character, sound pressure levels in the frequency region of 31.5 to 125 Hz were increased during data processing. Each sound was recorded for approximately 2 minutes (binaurally and monaurally) and these recordings were used to prepare the experimental sounds, which were 5 seconds and 10 minutes. Care was taken to prepare the 5-seconds sounds so that they were representative of the 10-minute sounds. Each sound was reproduced at 3 different levels (naturally occurring level at the recording time (0 dB), 6 dB below (-6 dB) and 6 dB above (+6 dB)). For ventilation sound the low-frequency-boosted version is referred to as the natural level. The equivalent A-weighted sound pressure levels (L_{Aeq}) of 10 minute sounds ranged from 52 to 59 dB while 5 second sounds ranged from 51 to 55 dB (natural level).

2.3 Exposure room and playback setup

The experiments were carried out in a room (l=8.10 m, w=6.96 m, h=3.05 m), which was partly furnished as a living room with a two-person sofa, two armchairs, a small table, and some plants. Figure 1 shows the listening test set up. The sound pressure level (SPL) of the background noise (including the ventilation and cooling system) was below the hearing threshold (ISO 226; 2003) for every 1/3 octave frequency band between 20 Hz and 12.5 kHz.

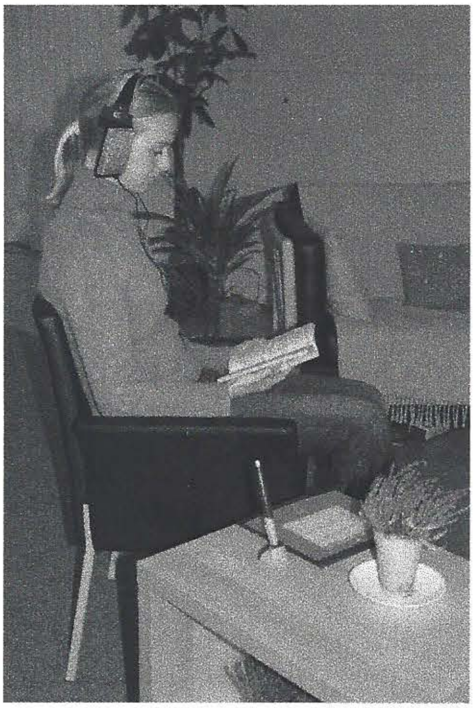

Figure 1: Listening test setup.

The monophonic recordings were presented through a loudspeaker system (Genelec 1031A/1094A) (technique ML), which was hidden behind a curtain, and the binaural recordings were presented through either circum-aural headphones (Beyerdynamics DT 990) (technique BH1) or headphones that were completely open and free of the ear (AKG K 1000) (technique BH2). In technique BH2, due to limitations (harmonic distortion during the playback of low-frequency sound) of the open headphone, it was necessary to play back the sound in a different way than normal binaural playback. The low-frequency part (lower than 100 Hz) was reproduced through the loudspeaker and the rest through the open headphone, so the subjects were fully exposed to the low-frequency sound field without loosing the spatial perception connected to the binaural technique.

2.4 Evaluation methods

In Method I, each group rated either annoyance, loudness or unpleasantness by answering the question: "How XX did you find the sound?" (XX was replaced by annoying, loud and unpleasant for the three different groups). The answers were given on an electronic tablet with a 100 mm horizontal scale with the anchor points "not at all XX" and "very XX". Degree of annoyance, loudness or unpleasantness was measured in mm and automatically stored on a computer after each exposure.

In Method II (paired comparisons) the subjects made forced-choice paired comparisons of annoyance, loudness or unpleasantness (depending on the group) of sounds. The two sounds in a pair were presented with a 1 second pause in between. The question was: "Which of the sounds were you more annoyed by?" or "Which of the sounds did you find louder?" or "Which of the sounds did you find more unpleasant?" The answers were given on an electronic tablet after each exposure, where one of two alternatives had to be chosen. The sounds in a pair were either from the same technique or from different techniques. Only ML and BH2 techniques were used in this session, since only these would allow comparisons across techniques without the need of taking the headphones on and off between the two sounds in a pair.

2.5 Experimental design and procedure

For Method I the study had for each group a 3 (sounds) $\times$ 3 (levels) $\times$ 3 (techniques) $\times$ 2 (durations) factorial design with repeated measures. The 10 minute stimuli were given on separate days with one technique per day, and subjects were asked to choose a book out of 5 alternatives and read it during the test. The 5 second stimuli were given on one day, and in order to allow an evaluation of the subjects' reliability all stimuli appeared twice. The order of techniques (ML, BH1 and BH2) was balanced between subjects (same order for 10 minute and 5 second experiments). The order of stimuli was randomized for each subject, technique and duration.

In Method II, 18 stimuli were included (2 techniques $\times$ 3 sounds $\times$ 3 levels). The pairs were taken from a half matrix design that excludes identical and reverse pairs, thus giving a total of 153 pairs ($n \times (n-1)/2$, $n=18$). The order of the pairs was randomized. With the given design, for comparisons within the same technique, each sound/level combination occurred once with any other sound/level combination, and for these the order of the two combinations was random. For across techniques comparisons, each sound/level combination occurred twice with any other sound/level combination. The first time the order of the techniques was random, while it was reversed the second time.

On a separate day before the experiment (preparation day), subjects underwent an audiometric test and filled in the noise sensitivity questionnaire. Each subject took part in sessions on 5 separate days (with a minimum of 48 hours in-between) and always at the same time of the day. During all sessions subjects were given breaks at regular intervals, in order to avoid tiredness. For each group half of the subjects completed Method I-5 second and Method II on their first experimental day, while the other half started with the three days of Method I-10 minute, see Table 1.

Prior to each method and technique, subjects were given written and verbal instructions, and they listened to 5 seconds of each sound in all levels in a random order. They also underwent a learning session in order to get familiar with the test method. The subjects were instructed to remain seated in the same position throughout the test. They were also informed that during the test they would be monitored by the operator (by mean of intercom and camera). Subjects were instructed to give their immediate response.

Table 1: The Experimental schedule for each of the three 18-subject groups (annoyance, loudness, unpleasantness).

Day	9 subjects	9 subjects
Prep. day	Audiometry Questionnaire	Audiometry Questionnaire
1. Day	Method I-5 s Method II	Method I-10 min
2. Day	Method I-10 min	Method I-10 min
3. Day	Method I-10 min	Method I-10 min
4. Day	Method I-10 min	Method I-5 s Method II
TOTAL	Preparation=35 min Experiment=315 min	Preparation=35 min Experiment=315 min

3 Results

3.1 Method I

In order to evaluate the influence of duration, technique, sound, level, attribute as well as interactions between these, a 5-way repeated measures analysis of variance (ANOVA) with 4 within subject factors (duration, technique, sound, level) and 1 between subject factor (attribute) was performed. The degrees of freedom of the corresponding F tests were corrected according to

Greenhouse-Geisser when sphericity was violated. The statistical analysis were carried out using SPSS. All tests were two-tailed, and a p-value below 0.05 was considered statistically significant (mean difference is abbreviated as MD and 95% confidence intervals are given in brackets). The main effects and significant interactions from the analysis of variance are shown in Table 2.

Table 2: The main effects and significant interactions from the analysis of variance (T=Technique, S=sound, L=Level, D=Duration, A=Attribute).

	df1;df2	F	p
T	2;102	0.704	0.497
D	1;51	1.409	0.241
S	1.48;75.25	40.903	0.000
L	1.20;61.03	289.392	0.000
S*D	1.60;81.84	29.437	0.000
S*T	4;204	14.307	0.000
L*D	1.58;80.80	19.561	0.000
S*L	4;204	15.565	0.000
S*A	2.95;75.25	4.141	0.004
L*A	2.39;61.03	4.955	0.007
T*L*A	8;204	2.133	0.034
T*S*L	8;408	2.097	0.035
S*L*A	8;204	2.968	0.004
D*S*A	3.21;81.84	8.521	0.000
D*S*L*A	8;204	1.998	0.048

No significant main effects of technique and duration on ratings were found.

A significant main effect of sound was found. Restaurant sound was significantly different from traffic (MD=11.13 [8.2;14]) and ventilation sounds (MD=9.18 [5.1;13.3]). Traffic and ventilation sounds were not significantly different.

A significant main effect of level was found. All levels were significantly different from each other (0 dB versus -6 dB: MD=9.7 [8.1;11.3]; +6 dB versus 0 dB: MD=11.6 [9.9;13.4]; +6 dB versus -6 dB: MD=21.3 [18.4;24.3]).

A significant two-way interaction was found between sound and duration. Traffic sound and in particular ventilation sound was rated higher with 5 second exposures than with 10 minutes exposure, while the reverse was observed for restaurant sound (Figure 2).

A significant two-way interaction was detected between sound and technique. The BH1 and BH2 techniques gave comparable ratings for all sounds. With the ML technique, restaurant sound was rated lower, traffic and ventilation sounds higher (Figure 3).

A significant two-way interaction was found between level and duration. The ratings increased more steeply with level for the 5 second exposures than for the 10 minute exposures (Figure 4).

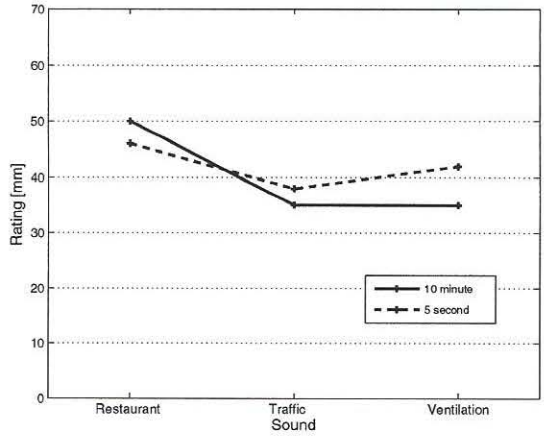

Figure 2: Mean rating as a function of sound and duration.

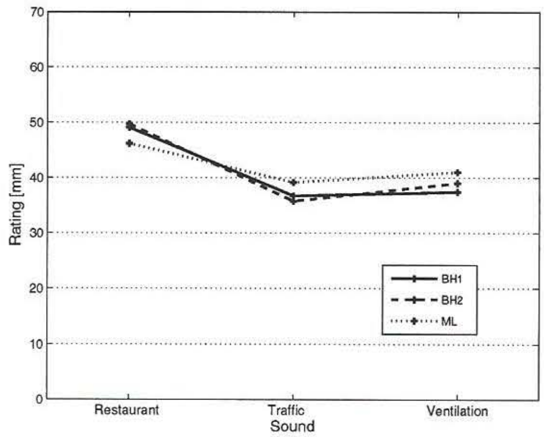

Figure 3: Mean rating as a function of sound and technique.

A significant two-way interaction was found between level and sound. The ratings increased more steeply with level for the ventilation sound than for restaurant and traffic sound (Figure 5).

A significant two-way interaction was detected between sound and attribute. A complicated pattern with no simple trends was seen (Figure 6).

A significant two-way interaction was also found between level and attribute. The ratings increased more steeply with level for loudness than for annoyance and unpleasantness (Figure 7).

3.2 Method II

For each independent group (N=18) the individual paired comparison matrices were pooled across subjects, resulting in the cumulative preference matrix. In this matrix

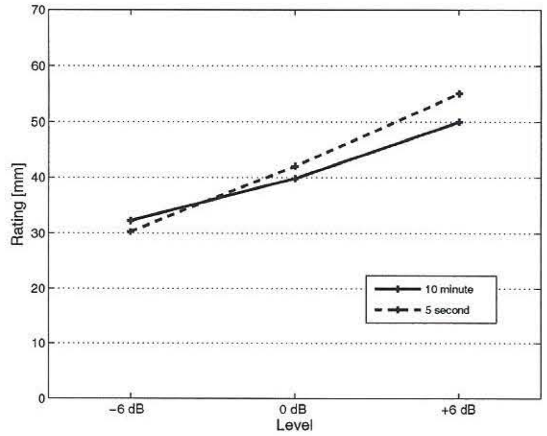

Figure 4: Mean rating as a function of level and duration.

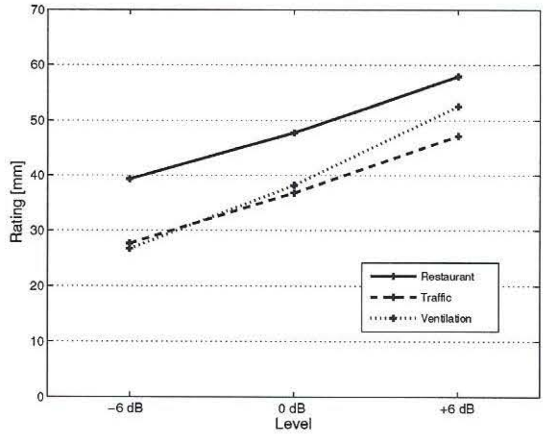

Figure 5: Mean rating as a function of level and sound.

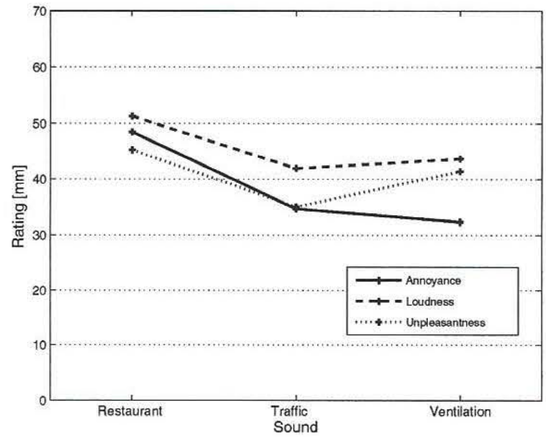

Figure 6: Mean rating as a function of sound and attribute.

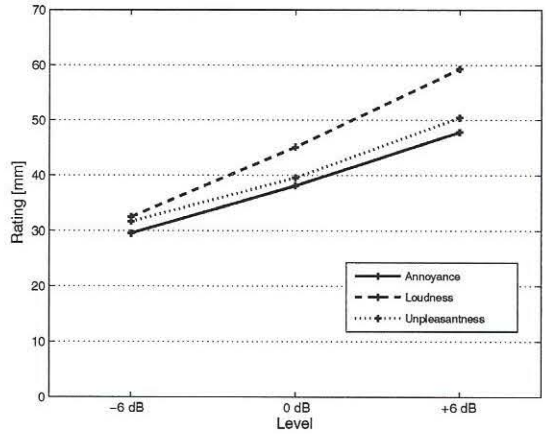

Figure 7: Mean rating as a function of level and attribute.

each entry specifies the absolute frequency with which the sound identified by the row of the table was judged as more annoying/loud/unpleasant than the sound identified by the column of the table. Stochastic transitivity checks were performed for each independent group [3]. The data did not fulfill the restrictions for the moderate and strong stochastic transitivity which are a prerequisite for a ratio scale [4]. Therefore the data were evaluated with respect to weak stochastic transitivity, a prerequisite for an ordinal representation of the data. For each cumulative matrix all the columns of each row were summed. This yields how many times a sound was preferred over the other sounds in the test. The result allows to determine the relative order (ranking) of the 18 sitimuli (Table 3; 4th, 7th, 10th column). In each group the rank order of the 9 sounds which were played back through open headphone were compared with the 9 sounds which were played back through loudspeaker using a Mann-Whitney test. The results did not show any significant difference between the two techniques for any of the independent groups.

The correlation between rank orders of the three attributes were calculated (Table 3; 4th, 7th, 10th column) and showed high correlation (A–L=0.961, A–U=0.930, L-U=0.926).

Furthermore, in order to be able to compare two different psychometric method (direct scaling and forced choice paired comparisons) the data for the same sounds from Method I were also ranked. Table 3 includes the relative order of the sounds both from Method I and II for each independent group.

The correlations between these 9 rankings were calculated. From the correlation coefficients, distance measures were derived and this new matrix was visualized by a multi dimensional solution (MDS) algorithm in two dimensions (see Figure 8). Scales that are close on the plot have high correlation, whereas scales that are more distant are less correlated.

Table 3: The relative order of the sounds for each independent group for Method I and II (sc=Method I (scale), pc=Method II (paired comparisons); l=10 min, s=5 s).

Sound	A			L			U		
	sc		pc	sc		pc	sc		pc
	l	s	s	l	s	s	l	s	s
R-BH2(-6)	15	10	7	8	6	6	13	4	4
R-BH2(0)	16	16	12	14	12	12	16	10	10
R-BH2(+6)	17	18	18	17	18	18	17	16	16
T-BH2(-6)	2	3	2	2	2	5	2	1	2
T-BH2(0)	6	8	9	5	7	8	5	5	9
T-BH2(+6)	10	13	14	11	13	16	9	11	15
V-BH2(-6)	1	1	4	1	1	4	3	9	6
V-BH2(0)	3	6	10	6	8	11	6	13	12
V-BH2(+6)	8	15	17	16	15	17	14	17	18
R-ML (-6)	11	7	5	9	5	2	10	3	3
R-ML (0)	14	11	11	12	11	10	12	6	8
R-ML (+6)	18	17	16	15	17	15	18	14	13
T-ML (-6)	5	4	1	4	4	3	4	2	1
T-ML (0)	9	9	6	7	10	7	8	8	7
T-ML (+6)	13	12	13	13	14	13	11	15	14
V-ML (-6)	4	2	3	3	3	1	1	7	5
V-ML (0)	7	5	8	10	9	9	7	12	11
V-ML (+6)	12	14	15	18	16	14	15	18	17

4 Conclusion

The results from direct scaling and paired comparison in general supported each other. None of the methods showed significant main effects of recording/playback techniques on psycho-acoustic attributes. Results from direct scaling, however, showed significant interactions between recording/playback techniques and sounds. For all sounds the two binaural techniques had good agreement with each other, while the monophonic technique gave lower ratings for the restaurant sound and higher for the traffic and ventilation sounds.

5 Acknowledgements

The project was sponsored by the Danish Technical Research Council. The authors gratefully acknowledge the technical assistance of Søren Krarup Olesen, Christian Sejer Pedersen and Steffen Pedersen.

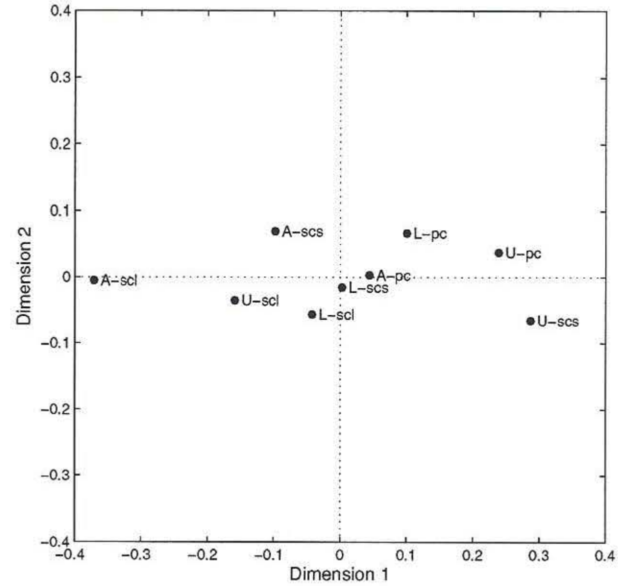

Figure 8: Classical multi dimensional solution of the correlation distance for Method I and II data.

References

[1] F. Christensen, C. B. Jensen, H. Møller, 'The design of VALDEMAR - an artificial head for binaural recording purpose'. *Proceedings of Audio Engineering Society's 109th Convention* 2000, Los Angeles, California, pp. 1-8 (2000)

[2] U. Landström, S. Arlinger, S. Hygge, Ö. Johansson, A. Kjellberg, K. Persson Waye: *'Annoying noise. Review for criteria documentation (in Swedish). Arbete och Hälsa'*, Research Report No. 1999:7, National Institute for Working Life, Solna, Sweden, July 1999.

[3] R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, package:'eba', URL http://www.R-project.org.

[4] Tversky, A., 'Intransitivity of preferences', *Psychological Review*, Vol.76. No.1. pp. 31-48, (1969)

[5] N. D. Weinstein, 'Individual differences in reactions to noise: A longitudinal study in a college dormitory', *J. Appl. Psychol*, Vol. 63. pp. 458-466. (1978)